Distinct molecular signatures of clinical clusters in people with type 2 diabetes

Slieker, Roderick C.; Donnelly, Louise A.; Fitipaldi, Hugo; Bouland, Gerard A.; Giordano, Giuseppe N.; Åkerlund, Mikael

Published in:
Diabetes

DOI:
10.2337/db20-1281

Publication date:
2021

Document Version
Peer reviewed version

Distinct molecular signatures of clinical clusters in people with type 2 diabetes: an IMI-RHAPSODY study

Roderick C Slieker, PhD[1,2,*], Louise A Donnelly, PhD[3,*], Hugo Fitipaldi, MSc[4], Gerard A Bouland, MSc[1], Giuseppe N. Giordano, PhD[4], Mikael Åkerlund, PhD[4], Mathias J. Gerl, PhD[5], Emma Ahlqvist, PhD[4], Ashfaq Ali, PhD[6], Iulian Dragan, MSc[7], Petra Elders, PhD[10], Andreas Festa, MD[8,9], Michael K. Hansen, PhD[11], Amber A van der Heijden, PhD[10], Dina Mansour Aly, MSc[4], Min Kim, PhD[6,12], Dmitry Kuznetsov, PhD[7], Florence Mehl, PhD[7], Christian Klose, PhD[5], Prof Kai Simons, PhD[5], Imre Pavo, MD[8], Timothy J. Pullen, PhD[13,14], Tommi Suvitaival, PhD[6], Asger Wretlind, MSc[6], Prof Peter Rossing, MD[6,15], Prof Valeriya Lyssenko, PhD[16,17], Cristina Legido Quigley, PhD[6,11], Prof Leif Groop, PhD[4,18], Prof Bernard Thorens, PhD[19], Prof Paul W Franks, PhD[4,20], Mark Ibberson, PhD[7], Prof Guy A Rutter, PhD[13,21], Prof Joline WJ Beulens, PhD[2,22], Leen M 't Hart, PhD[1,2,23,*,†], Prof Ewan R Pearson, MD[3,*,††]

[1] Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, the Netherlands

[2] Department of Epidemiology and Biostatistics, Amsterdam Public Health Institute, Amsterdam UMC, location VUMC, Amsterdam, the Netherlands

[3] Population Health & Genomics, School of Medicine, University of Dundee, Dundee, UK.

[4] Genetic and Molecular Epidemiology Unit, Lund University Diabetes Centre, Department of Clinical Sciences, CRC, Lund University, SUS, Malmö, Sweden

[5] Lipotype GmbH Dresden Germany

[6] Steno Diabetes Center Copenhagen, Gentofte, Denmark

[7] Vital-IT Group, SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland

[8] Eli Lilly Regional Operations GmbH, Vienna, Austria.

[9] 1st Medical Department, LK Stockerau, Niederösterreich, Austria.

[10] Department of General Practice and Elderly Care Medicine, Amsterdam Public Health Research Institute, Amsterdam UMC–location VUmc, Amsterdam, the Netherlands
[11] Cardiovascular and Metabolic Disease Research, Janssen Research & Development, Spring House, Pennsylvania, U.S.A.
[12] Institute of Pharmaceutical Science, Faculty of Life Sciences and Medicines, King's College London, London, United Kingdom
[13] Department of Diabetes, Guy's Campus King's College London, U.K.
[14] Section of Cell Biology and Functional Genomics, Division of Diabetes, Endocrinology and Metabolism, Department of Metabolism, Digestion and Reproduction, Imperial College London, London, U.K.
[15] Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
[16] Department of Clinical Science, Center for Diabetes Research, University of Bergen, Bergen, Norway
[17] Genomics, Diabetes and Endocrinology Unit, Department of Clinical Sciences Malmö, Lund University Diabetes Centre, Skåne University Hospital, Malmö, Swede
[18] Finnish Institute of Molecular Medicine, Helsinki University, Helsinki, Finland
[19] Center for Integrative Genomics, University of Lausanne, CH-1015, Lausanne, Switzerland
[20] Department of Nutrition, Harvard School of Public Health, Boston, MA, USA
[21] Lee Kong Chian School of Medicine, Nan Yang Technological University, Singapore
[22] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands
[23] Department of Biomedical Data Sciences, Section Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands
* Contributed equally

Corresponding authors
[†] Leen 't Hart, PhD
Einthovenweg 20
2333ZC Leiden
The Netherlands
Telephone: +31 71 5269796
E-mail: lmthart@lumc.nl

[††] Prof Ewan Pearson, MD

Ninewells Hospital

Dundee

DD1 9SY

United Kingdom

Telephone: +44 1382 383387

E-mail: E.Z.Pearson@dundee.ac.uk

Running title: Molecular signatures of clusters in diabetes

Tweet

Distinct molecular signatures of clinical clusters in people with type 2 diabetes: an IMI-RHAPSODY study. @IMI_Rhapsody@imi_ju@rcslieker@ezpearson@lmthart1#Clusters#type2diabetes#RHAPSODY#Lipidomics#Metabolomics#Proteomics

ABSTRACT

Type 2 diabetes is a multifactorial disease with multiple underlying aetiologies. To address this heterogeneity a previous study clustered people with diabetes into five diabetes subtypes. The aim of the current study is to investigate the aetiology of these clusters by comparing their molecular signatures. In three independent cohorts, in total 15,940 individuals were clustered based on five clinical characteristics. In a subset, genetic- (N=12828), metabolomic- (N=2945), lipidomic- (N=2593) and proteomic (N=1170) data were obtained in plasma. In each datatype each cluster was compared with the other four clusters as the reference. The insulin resistant cluster showed the most distinct molecular signature, with higher BCAAs, DAG and TAG levels and aberrant protein levels in plasma enriched for proteins in the intracellular PI3K/Akt pathway. The obese cluster showed higher cytokines. A subset of the mild diabetes cluster with high HDL showed the most beneficial molecular profile with opposite effects to those seen in the insulin resistant cluster. This study showed that clustering people with type 2 diabetes can identify underlying molecular mechanisms related to pancreatic islets, liver, and adipose tissue metabolism. This provides novel biological insights into the diverse aetiological processes that would not be evident when type 2 diabetes is viewed as a homogeneous disease.

Type 2 diabetes is a multifactorial disease with multiple underlying aetiologies.(1; 2) In an attempt to address this heterogeneity, a recent study stratified people with any form of diabetes into five clusters based on six clinical variables, i.e. age, glutamate decarboxylase (GAD) antibodies, BMI, HbA1c, insulin resistance (HOMA2-IR) and β-cell function estimates (HOMA2-B).(3) Based on this work, we clustered and cross-validated individuals into five clusters in three large cohorts based on age, BMI, random or fasting c-peptide, HbA1c and HDL, largely reproducing the ANDIS clusters using more readily measured clinical variables.(4)

The original and subsequent papers have shown that people in different clusters had different risks for a number of diabetes related outcomes.(3; 5-7) The autoimmunity and insulin deficient cluster were defined by high HbA1c at diagnosis, had ketoacidosis and retinopathy(7) more often and progressed more rapidly onto insulin compared to the other clusters.(3) The insulin resistant cluster showed a higher frequency of non-alcoholic fatty liver disease and people in this group were at increased risk of developing chronic kidney disease.(3) The differences in progression and characteristics of the different clusters suggest that these groups represent different underlying aetiologies. For example, differences in genotype frequency across clusters based on candidate loci were observed and this was further illustrated in a follow-up study where it was shown that individuals in different clusters have differences in portioned polygenic risk scores for diabetes-related outcomes.(3; 8)

A systematic deconvolution of the different etiological processes underlying the clusters is currently lacking. To address this, we investigate each cluster's molecular signature using metabolomics, lipidomics, proteomics, and genomics to better understand the underlying aetiological processes representative of patients with diabetes in that cluster.

RESEARCH DESIGN AND METHODS

Cohort descriptions

Data from 15,940 individuals from three cohorts, DCS (Netherlands), GoDARTS (Scotland) and ANDIS (Sweden) were used in this cross-sectional study. Inclusion criteria for RHAPSODY were age of diagnosis was ≥35 years, clinical data available within 2 years after diagnosis, GAD negative, no missing data in one of the five for clustering used clinical measures and the presence of GWAS data. Individuals were clustered using k-means clustering based on five clinical characteristics age at sampling, BMI, HbA1c, HDL and C-peptide. Of note, C-peptide was included in the clustering as proxy of insulin resistance, while HDL has previously been recognized as risk factor for time to insulin requirement. Details on the cohorts and clustering have been described elsewhere.(4) Briefly, DCS is an open prospective cohort that started in 1998 comprised on over 14,000 individuals with type 2 diabetes from the northwest part of the Netherlands.(9) The Ethical Review Committee of the VU University Medical Center, Amsterdam has approved the study. People visit DCS annually as part of routine care. GoDARTS is a study comprising individuals with diabetes mellitus from the Tayside region of Scotland (N = 391,274; January 1996) that were added to the DARTS register.(10) The GoDARTS study was approved by the Tayside Medical Ethics Committee. Longitudinal retrospective and prospective anonymized data were collected, including data on prescribing, biochemistry, and clinical data. In ANDIS, people were recruited with incident diabetes within the Scania County, Sweden from January 2008 until November 2016.

Molecular measures

An overview of the sample selection procedure is given in Fig. S1a. Individuals were selected based on the shortest time between diagnosis date and sampling date without taking into account cluster assignment. Analysis of small charged molecule analytes (metabolomics, UHLPC-MS/MS) was performed in the largest set (N=2945), followed by lipidomics (N=2593, Lipotype lipidomics platform) and proteomics (N=1170, SomaScan® Platform- Somalogic). Of note, the smaller sets were selected from the larger set based upon the samples being collected closest to the time of diagnosis, so in the smallest set of 1170 GWAS, metabolomics, lipidomics and proteomics was available (Fig. S1a). Molecular measures were taken close to diagnosis (Table S2). Quality control was performed in a similar way for metabolomics, lipidomics and proteomics. A participant's data was excluded if their profile was a strong outlier

based on principal components analysis and the data of the individual measurements was clearly distinct from the other samples.

Genetic data

In DCS, genetic data were generated using the Illumina HumanCoreExome array. In GoDARTS genetic data were generated using the Affymetrix Genome-Wide Human SNP Array 6.0 and the Illumina HumanOmniExpress Array. ANDIS was genotyped InfiniumCoreExome-24v1-1 BeadChip arrays (Illumina, San Diego, CA, USA), at Lund University Diabetes Centre, Malmö, Sweden. Samples were excluded for ambiguous gender, call rate < 95%, and any duplicate or related individuals (pi_hat ≥ 0.2). SNP were excluded for monomorphic SNPs, SNPs with MAF < 0.05, and SNPs with missingness rate > 0.05. Differences in diabetes-related genetic risk were based on 403 relatively independent diabetes associated SNPs identified in a recent large GWAS meta-analysis.(11) Genetic data were imputed using the Michigan Server against the reference panel Human Reference Consortium R1.1 using default settings, i.e. phasing with Eagle v2.3 and population of European descent.(12) SNPs with minor allele frequency below 5% were discarded from the analyses leaving 394 SNPs across the three studies.

Metabolomics

Fifteen small charged molecules were measured in plasma using targeted UHLPC-MS/MS (Steno Diabetes Center, Copenhagen, Denmark).(13) In DCS, 1267 individuals were included for metabolomics measurements. All passed QC and 1230 individuals overlapped with the cluster data. In GoDARTS, 898 individuals were included in the analysis, one failed QC and of the 897 remaining individuals, 894 overlapped with the cluster data. In ANDIS, 896 individuals were included in the analysis, four failed QC and of the 892 remaining samples, 821 overlapped with the cluster data.

Lipidomics

614 plasma lipids common to the three cohorts were determined using a QExactive mass spectrometer (Thermo Scientific) equipped with a TriVersa NanoMate ion source (Advion Biosciences) on the Lipotype lipidomics platform (Lipotype, Dresden, Germany).(14) Samples were divided into analytical batches of 84 samples each. Lipid identification was performed on unprocessed mass spectra files using LipotypeXplorer.(15) Only lipid identifications with a signal-to-noise ratio >5, and a signal intensity 5-fold higher than in corresponding blank

samples were considered for further data analysis. Batch correction was applied using eight reference samples per 96-well. Amounts were also corrected for analytical drift if the p-value of the slope was below 0.05 with an R^2 greater than 0.75 and the relative drift was above 5%. In DCS, 900 individuals were included for lipidomics measurements, all passed QC and 877 overlapped with the cluster data. In GoDARTS, 898 individuals were included in the analysis, one failed QC and of the 897 remaining samples, 894 overlapped with one of the clusters. In ANDIS, 896 individuals were included in the analysis, five failed QC and of the 891 remaining samples, 820 overlapped with one of the clusters. Lipid nomenclature is used as described previously and SwissLipids database identifiers are provided (Table S1).(16) After quality control 162 lipid species were used in this study. The median coefficient of subspecies variation of the 162 lipids used as accessed by reference samples was 9.49% across all three cohorts.

Protein measurements

Protein levels (1195 proteins) in plasma were measured on the SomaLogic SOMAscan platform (Boulder, Colorado, USA) in 600 individuals each for both DCS and GoDARTS. Individuals were removed if they were strong outliers based on a principal component analysis. In DCS, 600 individuals were included for proteomics measurements, 11 failed QC and 573 overlapped with one of the cluster data. In GoDARTS, 600 individuals were included in the analysis, one failed QC and of the 599 remaining samples, 597 overlapped with one of the clusters.

Statistical analysis

Molecular data were log-transformed and z-scaled before analysis on a federated node system. Each of the cohorts' data were stored on a local node using Opal, an open source data warehouse (Open Source Software for BioBanks, OBiBa). A central node responsible for federated node access, user administration and software deployment was set up at SIB. Clinical and molecular data were harmonized according to the CDISC Study Data Tabulation Model (www.cdisc.org).

To identify molecular measures specific for a cluster, a generalized linear model was used to test each of the molecular measures in each cluster, where cluster *i* was compared against reference group *j*, where *j* was a combined group of the other clusters. Effect sizes represent change per log standard deviation of the tested molecular measure. Genetic data were not transformed and represent change in allele frequency. For example, cluster 1 was compared to clusters 2-5 combined, cluster 2 to clusters 1,3,4,5. Main results presented are based on an unadjusted model (log and z-scaled). Next, as an exploratory sensitivity analysis, models were adjusted for the extreme characteristic of a cluster to investigate whether the observed effect

was independent of the extreme characteristic. This was only done for those clusters that had extreme characteristics. Models were run on each of the cohorts separately and meta-analysed using the R-package *meta.*(17) Meta-analysed *P*-values were adjusted using the Benjamini-Hochberg procedure and a false discovery rate-adjusted (FDR) *P*-value below 0.05 was considered significant.

Partitioned polygenic risk scores (pPRSs) were obtained from Udler et al.(18). In each individual cohort, dosages of SNPs were multiplied with the scores for each cluster, which resulted in a risk score per individual for each of the five clusters beta cell (30 SNPs), proinsulin (7 SNPs), obesity (5 SNPs), lipodystrophy (20 SNPs) and liver (5 SNPs). Differences in pPRSs were tested with a linear model for one cluster with the other clusters as the reference group. Next, results from the three cohorts were meta-analysed using the metagen function from the meta package. P-values were Bonferroni adjusted and considered significant at $P_{adj} < 0.05$.

Pathway enrichment on the proteomics was performed based on KEGG pathways using the R-package of STRINGdb (1.24.0). The entire Somalogic set (1195 proteins) was used as the background set. *P*-values of enriched pathways were adjusted using the Benjamini-Hochberg procedure and an FDR-adjusted *P*-value below 0.05 was considered significant.

Effect sizes of proteins associated with eGFR and incident cardiovascular disease (CVD) were obtained from Yang et al. (2020).(19) Up- and downregulated proteins in each of the clusters ($P_{FDR}<0.05$) were selected and compared to the from Yang et al. obtained 1) correlation coefficient of protein levels and eGFR and 2) hazard ratios from the Cox Proportional Hazard models for cardiovascular disease in non-CKD individuals.(19)

Analyses were performed using R statistics (version 3.6.2). Figures were produced using the R-package *ggplot2* (v3.3.0) and *omicCircos* (v1.22.0).

Data and resource availability statement

The datasets generated during and/or analysed during the current study are not publicly available, but are available from the corresponding author upon reasonable request.

RESULTS

In this cross-sectional study, 15,940 individuals from three cohorts were included described previously.(4) As described, we reproduced the original ANDIS SIDD, SIRD and MOD clusters; and refined the MARD cluster into two, a subset with high HDL (MDH) and one without any particular defining features (MD). The characteristics of the clusters and those of the individuals used for molecular characterization (genetic data, metabolites, lipids, and proteins) are given in **Table S2** and **Table S3**.

Severe insulin-deficient cluster (SIDD)

For SIDD, no differences were observed in allele frequency of known type 2 diabetes *loci* compared to the other clusters (**Table S4**), nor in the pPRS (**Fig. S2**). Two metabolites, tyrosine (**Fig. 1a, Fig. 1b**) and asymmetric/symmetric dimethylarginine (SDMA/ADMA, **Fig. 1c**, **Table S5**) were significantly lower in SIDD versus all other clusters. The effect sizes attenuated slightly after adjustment for the primary variable HbA1c that defined the SIDD cluster (**Fig. S3b**, **Table S5**). Of the lipids, eight were downregulated and one upregulated. Three of the eight downregulated lipids were of the sphingomyelin class, four lipids of the phosphatidylcholine class and one cholesterol ester (**Fig. 2a**, **Table S6**). The sole upregulated lipid was the cholesterol ester (CE 20:2;0). Seven out of nine lipids remained significant after adjustment (**Fig. S3c**, **Table S6**). Finally, eight proteins were differentially expressed with four up- and four downregulated (**Fig. S4a-d**), where the effect sizes remained similar after adjustment (**Fig. S3d**, **Table S7**).

Severe insulin resistance cluster (SIRD)

The SIRD cluster was characterized by a strong and distinct molecular signature of insulin resistance. The pPRS for beta-cell function and proinsulin (18) were decreased in the SIRD cluster relative to other clusters (beta cell, β[95%CI]=1.41[-2.21 – -0.62]; proinsulin, β[95%CI]=-0.28[-0.41 – -0.15], **Fig. S2**), indicating genetically higher beta-cell function in the SIRD group. Five diabetes-associated SNPs all showed a *lower* risk allele frequency. The top SNP (rs3802177-A) of SIRD mapped to the protective allele in *SLC30A8* (**Table S4**, **Table 1**). In a sensitivity analysis, only the *SLC30A8* variant remained significant after adjustment for C-peptide (**Fig. S3a**, **Table S4**, **Table 1**). The SIRD cluster showed eight upregulated metabolites, including four amino acids, i.e. tyrosine, leucine, isoleucine, and phenylalanine (**Fig. 1a**, **Fig.**

S5a-b). Two were metabolites of the amino acid L-tryptophan, i.e. L-kynurenine and indoxyl sulphate. Adjustment for C-peptide attenuated the effect (**Fig. S3b**, **Table S5**).

Eighty-nine lipids were changed in SIRD, with 45 (50.6%) upregulated and 44 downregulated (49.4%, **Fig. 2a**). Of the 45 upregulated lipids, 43 were in the di- and triacylglycerol class with TAG 51:3;0 as the strongest associating lipid (**Fig. 2b**, **Table S6**), while the remaining two upregulated lipids were phosphatidylcholines containing the omega-3 fatty acid docosahexaenoic acid (22:6;0, PC 18:0;0_22:6;0, PC 16:0;0_22:6;0). Of the 44 downregulated lipids, the most represented were the phosphatidylcholines class (27 lipids, 61.4%), especially with the ether phosphatidylcholines (38.6%), with PC O-16:0;0/18:1;0 being the strongest downregulated lipid (**Fig. 2a, Fig. 2c, Table S6**). Also, most ether phosphatidylethanolamines are (four lipids, 9.1%) and sphingomyelin species were downregulated (7 lipids, 15.9%). The changes in lipids seemed to be dependent on the high C-peptide levels with effect sizes of DAG and TAGs close to zero after adjustment for the latter (**Fig. S3c**, **Table S6**).

Out of the 1195 plasma proteins investigated, 367 proteins were differentially expressed, with 158 proteins downregulated and 209 upregulated. Several top proteins were upregulated independent of C-peptide levels, including two metalloproteinases, matrix metalloproteinase-7 (MMP-7) and Macrophage metalloelastase (MMP-12), and MIC-1 (**Table S7**). Metalloproteinases are associated with multiple physiological processes, but also with atherosclerosis and diabetes-related nephropathy.(20; 21) MIC-1 (GDF-15) is known to be associated with insulin resistance.(22) The identified proteins showed a strong enrichment in pathways, including *Cytokine-cytokine receptor interaction* (50 proteins, $P_{FDR}=8.69\cdot10^{-56}$), *Chemokine signalling pathway* (26 proteins, $P_{FDR}=1.81\cdot10^{-34}$), *Axon guidance* (26 proteins, $P_{FDR}=3.55\cdot10^{-34}$) and *PI3K-Akt signalling pathway* (29 proteins, $P_{FDR}=1.05\cdot10^{-29}$). There was a significant reduction in 3-phosphoinositide-dependent protein kinase-1 (PDPK1, Fig. 3a, Fig. 3c), which, when activated by insulin, activates Akt/PKB and increases glucose uptake via GLUT4.(23) Plasma Akt itself was also decreased in SIRD (**Fig. 3a**). Insulin tended to be higher in SIRD although not significant (**Fig. 3b**), while the insulin receptor was significantly upregulated (**Fig. 3a**). In the downstream signalling cascade of the PI3K-Akt pathway, PDPK1 (**Fig. 3c**), RAC1, AMPK, HSP90, 14-3-3 and p53 were differentially expressed (**Fig. S5c-i**). Of note, the proteins associated with SIRD were only modestly driven by C-peptide levels (**Fig. S3d**).

Next, we overlapped identified proteins with those previously associated with eGFR and incident CVD.(19) Proteins upregulated in SIRD, were previously associated with lower

eGFR levels, including Cystatin C (ρ= -0.74, P= $1.12 \cdot 10^{-163}$), Tumor Necrosis Factor receptor superfamily member 1A (TNF sR-I, ρ= -0.65, P= $2.51 \cdot 10^{-114}$) and Neuroblastoma suppressor of tumorigenicity 1 (DAN, ρ= -0.64, P= $2.29 \cdot 10^{-109}$, **Fig. S7a**). Conversely, proteins positively associated eGFR were downregulated including Epidermal growth factor receptor (ERBB1, ρ= 0.44, P= $1.96 \cdot 10^{-46}$) and Alpha-2-antiplasmin (ρ= 0.41, P= $1.42 \cdot 10^{-38}$). For incident cardiovascular disease (CVD), Angiopoietin-2 (HR=1.66, P=$2.20 \cdot 10^{-16}$) and MMP-12 (HR=1.65, P=$2.20 \cdot 10^{-16}$) were upregulated risk factors in SIRD, while ERBB1 (HR=0.59, P=$2.20 \cdot 10^{-16}$) was protective for CVD and downregulated in SIRD **(Fig S7b)**.

Mild Obesity-related Diabetes (MOD)

In MOD, the pPRS for obesity was significantly higher (β[95%CI]=0.51[0.34 – 0.68], **Fig. S2**) compared to other clusters. Individual diabetes-associated risk alleles associated with high BMI were also more frequent in MOD, that is FTO (rs1421085-C) and the MC4R locus (rs523288-T, **Table S4**, **Table 1**). Of note, both loci are also in the pPRS, although using different SNPs in LD. Naturally, adjustment for BMI attenuated the effect size for both SNPs (**Fig. S3a**, **Table S4**).

Isoleucine was the sole metabolite that was differentially upregulated in MOD (**Fig. 1a**, **Fig. S4a**, **Table S5**), and this difference was completely eliminated after adjustment for BMI. The lipid profile of the MOD cluster was largely similar to the SIRD cluster (**Fig. 2a, Table S6**). That is, in MOD, acyl phosphatidylethanolamine species were upregulated, but not the ether phosphatidylethanolamines. Cholesterol esters and phosphatidylcholine species containing the omega-3 fatty acids eicosapentaenoic acid (20:5;0) and docosahexaenoic acid (22:6;0) were downregulated, while these were upregulated or not significantly changed in the SIRD cluster. However, cholesterol esters and phosphatidylcholine species containing 20:3;0 fatty acids are upregulated in MOD, while downregulated or not significantly changed in the SIRD cluster. In total 61 lipids were affected of which 40 were upregulated. Amongst these the diacylglycerols (15%) and triacylglycerols (57.5%) were strongly enriched. Of the 21 downregulated lipids, the majority were phosphatidylcholines (61.9%). The effect size for diacylglycerol and triacylglycerol changes were strongly reduced after adjustment for BMI (**Fig. S3c, Table S6**). Interestingly, the largest effect size was seen in the TAGs with the lowest number of acyl chain carbons and double bonds (**Fig. S6a-b),** while the TAGs with more acyl chain carbons and double bonds were not significantly altered in MOD. In a previous study, saturated or monounsaturated TAGs were associated with an increased diabetes risk, including

TAG 46:1, TAG 48:0 and TAG 48:1 that were also significantly upregulated in the MOD cluster.(24)

Of the 1195 proteins, 261 were differentially expressed in MOD with the majority downregulated (158 proteins, 60.5%, **Table S7**). After adjustment for BMI, several remained significant, although their effect sizes were attenuated, including NCAM-120, DKK3 and CRDL1 (**Fig. S3d, Table S7**). DKK3 has been associated with increased adipogenesis in fat cells.(25) CRDL1 has been shown to be predictive of beta-cell function.(26) The role of NCAM-120 is largely unclear. The strongest enrichment was found for *Cytokine-cytokine receptor interaction* with 38 proteins (42.7%, $P_{FDR}=2.08\cdot10^{-43}$) overlapping (**Fig. S8**). The strongest upregulated proteins in this pathway were leptin (**Fig. S4b**), growth hormone receptor and Interleukin-1 receptor antagonist protein, while Interleukin-1 receptor type 1(IL-1 sRi) was downregulated. Adjustment for BMI influenced the effect size of several proteins, including leptin, FABP and CRP (**Fig. S3d, Table S7**). Finally, upregulated proteins identified in MOD were generally positively associated with eGFR and protective for CVD, including the growth hormone receptor (HR=0.62, $P=2.20\cdot10^{-16}$, **Fig. S7**).

Mild diabetes with high HDL

The MDH cluster showed a higher GRS relative to the other clusters for beta-cell function (β[95%CI]=0.61[0.33-0.38], **Fig. S2**). Among the diabetes-associated SNPs, a lower risk allele frequency was observed for a SNP near *LPL* (rs10096633-T, **Table S4**, **Table S1**). With respect to metabolite-, lipid- and peptide levels the MDH cluster showed opposite effects compared to the SIRD and MOD cluster. The amino acids that were upregulated in SIRD were generally downregulated in MDH (**Fig. 1a**, **Table S5**). Only the difference in isoleucine level was significant and phenylalanine borderline insignificant. In addition, taurine was significantly upregulated in MDH. After adjustment for HDL the effect sizes strongly attenuated (**Fig. S3b, Table S5**).

Out of the 162 lipids, 135 lipids were affected in MDH, with 52 downregulated and 83 upregulated (Table S6). Opposite to SIRD and MOD, diacylglycerols (13.5%), triacylglycerols (73.1%) and acyl phosphatidylethanolamines (9.6%) were downregulated in MDH, while phosphatidylcholines (65.1%) were upregulated, especially the ether phosphatidylcholines (PC O-, 25.6%, **Table S6**). The TAGs with a smaller number of acyl chain carbons and double bonds showed the lowest protein levels versus the other clusters, while the differences attenuated with increasing number of acyl chain carbons and double bonds (**Fig. S6a-b**). In addition, upregulation was seen for cholesterol esters (13.3%), sphingomyelins (10.8%) and all

ether phosphatidylethanolamines (9.6%), which point in the opposite direction in the SIRD cluster (**Table S6**). Adjustment for HDL strongly decreased the effect size for diacylglycerols and triacylglycerols (**Fig. S3c, Table S6**).

Out of the 1195 proteins, 270 proteins were differentially expressed in the MDH cluster (119 down, 151 upregulated). The effect size of the proteins changed very modestly after adjustment for HDL (**Fig. S3d**). The peptide profile of the MDH cluster was opposite of that of MOD (**Fig. S9**, r=-0.82). As such among the top proteins similar proteins were identified such as CRDL1 that remained significant after adjustment for HDL. The pathway enrichment resembled that of SIRD and MOD, with enrichment for *Cytokine-cytokine receptor interaction* (31 proteins, $P_{FDR}=7.04\cdot10^{-32}$), *Pathways in cancer* (22 proteins, $P_{FDR}=2.35\cdot10^{-24}$) and *PI3K-Akt signalling pathway* (22 proteins, $P_{FDR}=5.56\cdot10^{-23}$). In the latter, growth hormone receptor was downregulated, as well as insulin (**Fig. 3b, Table S7**). Effect sizes were generally not solely driven by increased HDL levels (**Fig. S3d, Table S7**). MDH-associated proteins in relation to eGFR showed a similar pattern to that of SIRD (**Fig. S7a-b)**, with proteins associated with lower eGFR being upregulated as well as proteins associated with higher risk for CVD, the latter including Follistatin-related protein 3 (HR=1.55, $P=2.20\cdot10^{-16}$) and HCC-1 (HR=1.54, $P=2.20\cdot10^{-16}$).

Mild diabetes

The MD cluster was generally less well-defined, with only one significant SNP and no significant GRSs, lipids or metabolites. There was a higher risk allele frequency (C-allele) in MD – opposite to that of MDH – compared to the other clusters near the *LPL* gene (rs10096633-T, **Table S4**). In contrast to the few signals for lipids or metabolites, 354 proteins were differentially expressed in the MD cluster, with the majority downregulated (209 proteins, 59.0%). Enrichment was found for *Axon guidance* (20 proteins, $P_{FDR}=1.12\cdot10^{-30}$), *Cytokine-cytokine receptor interaction* (25 proteins, $P_{FDR}=3.48\cdot10^{-25}$), *PI3K-Akt signalling pathway* (21 proteins, $P_{FDR}=4.28\cdot10^{-23}$). While similar pathways were found to be enriched compared to the SIRD cluster the effect sizes were correlated but reversed (r=-0.88, **Fig. S9**). In line with this, insulin and its receptor were significantly downregulated in MD. Finally, in MD upregulated proteins were generally associated with better eGFR levels and lower risk for CVD (**Fig. S7a-b**)

DISCUSSION

Based on five clinical variables, people with type 2 diabetes from three large European cohorts were assigned to five separate clusters. The molecular phenotyping of the clusters revealed that, in addition to differences in clinical characteristics, there were also profound differences in underlying molecular profiles which related to pancreatic islet biology (in SIDD), liver (in SIRD) and adipose tissue metabolism (in MOD and MDH).

The SIRD cluster was characterized by a molecular profile that fits with insulin resistance, i.e. upregulation of DAGs, BCAAs and insulin and downregulation of PI3K-Akt pathway-related proteins and phosphatidylcholines. The MOD cluster showed overlap with the SIRD cluster, but with a more pronounced molecular profile of obesity. Individuals in the MDH cluster showed the opposite effect of SIRD and MOD with, relative to the other clusters, low levels of TAG, DAG and BCAAs, but higher levels of ether phosphatidylcholines and phosphatidylethanolamines, sphingomyelins, and cholesterol esters. The results were in part, but not fully, driven by the identifying characteristic of the cluster, except for SIDD which showed consistent results after adjustment for HbA1c. For example, effect sizes of TAGs and DAGs in SIRD and MDH were influenced by adjustment for C-peptide and HDL, respectively. The lower frequency of diabetes-associated risk alleles could be explained by the fact that most diabetes SNPs are associated with reduced insulin-secretion. People in the SIRD cluster do not have diabetes because of lower insulin secretion but because of high insulin resistance (and consequent greater beta-cell function).

The SIDD cluster was characterized by greater insulin sensitivity and lower beta-cell function than the other clusters based on the clinical characteristics. SIDD is characterized by low tyrosine levels and (a)symmetric dimethylargine, CE 16:1;0, PC O-34;1 and PC O-34;2, compared to the other clusters; higher levels of these metabolites and lipids have been associated with higher type 2 diabetes risk.(27-30) Higher CE 16:1;0 has also been associated with higher fasting plasma glucose (FPG) and 2-hour post-loading glucose (2h-PLG).(31) Moreover, in SIDD, CRP was downregulated and this is in line with a previous report that CRP levels are generally higher in those with insulin resistance and not low secretion.(32)

The SIRD – and to some extent the MOD cluster – showed opposing metabolite, lipid and protein profiles compared to the MDH cluster (Fig. 4). The SIRD cluster was characterized by a molecular signature compatible with insulin resistance inside cells. In SIRD, the frequency of protective alleles was higher for HOMA-B-associated variants. Evidence was found for downregulation of insulin-mediated glucose uptake across the different omics levels, where for example higher levels of BCAA and DAG/TAG were observed. BCAAs have been shown to

be risk factors for developing incident type 2 diabetes in observational studies; their causal role has also been suggested.(33) Both BCAAs and DAG inhibit insulin receptor substrate 1 (IRS1).(34) DAGs activate PKC isoforms which inhibit PI3K activation by phosphorylating the inhibitory serine 307 of IRS1 instead of tyrosine.(34; 35) BCAAs target the intramuscular mammalian target of rapamycin/ribosomal protein S6 kinase beta-1 (mTOR/p70S6K) signalling pathway as shown in *in vitro* and rodent *in vivo* studies that also inhibits the PI3K/Akt pathway via IRS1 and IRS2 depending on the cell type.(34) Inhibition of PI3K/Akt reduces the GLUT4 translocation. In SIRD multiple proteins were downregulated in PI3K/Akt and the GLUT4 translocation pathway, including Akt, PDPK1, RAC1, while insulin was strongly upregulated.(36; 37) Furthermore, upregulation was seen in three ephrin family members (Ephrin A2,A2,A5). Inhibition of the ephrin receptors has been shown to enhance glucose-stimulated insulin secretion in mice.(38) Although these results might suggest changes in the insulin or glucose responsiveness of relevant metabolic tissues (e.g. muscle, liver or adipose), proteins were measured in plasma in the current study and, as such, are unlikely to reflect changes in intracellular signalling. Future studies will be needed to determine the tissue(s) of origin of these biomarkers and the mechanisms through which they are released. For example, tissue-specific knock-out of proteins identified in plasma in cell lines or model organisms might provide insight into both the role and tissue of origin. The higher BMI in individuals in the MOD cluster was in line with the higher allele frequency of variants associated with a higher BMI, i.e. variants near *FTO* and *MC4R*. Interestingly also variants near *TM6SF2* were associated with this cluster. *TM6SF2* is known to be associated with NASH.(39) The metabolic and lipid profile of MOD resembled that of SIRD. An interesting observation was that the number of acyl chain carbons and double bonds was associated with the effect size in some clusters in particular MOD and MDH. In MOD lipids with a higher number of acyl chain carbons and double bonds the effect size was much lower compared to those with lower numbers. These findings are in line with a previous publication that showed that TAGs with a lower number of acyl chain carbons and double bonds are elevated in T2D cases versus controls.(24) In addition, lipids that were associated with increased diabetes risk were generally saturated or monounsaturated fatty acids.(24) MOD was further characterized by upregulation of leptin, growth hormone receptor and multiple interleukins and IL-1Ra. People with a high BMI have high levels of leptin, which may be a marker of leptin resistance.(40) IL-1Ra is negatively correlated with quantitative insulin sensitivity check index (QUICKI), where higher levels associate with higher insulin resistance.(32)

The MDH cluster was the cluster with the most beneficial profile and had a molecular signature of insulin sensitivity. This cluster had high HDL levels, low BCAA levels, low DAGs, and high levels of ether phosphatidylcholines relative to the other clusters (**Fig. 4**). Regarding the peptide level, the effects were opposite of the MOD cluster. MDH cluster displayed high levels of anti-inflammatory fatty acids which have been associated with improved insulin sensitivity in animal studies(41-43).

In the study by Ahlqvist et al.(3) the SIRD cluster was associated with poorer renal function. In the current study we compare the identified proteins to proteins previously associated with eGFR levels and CVD-risk.(19) We show that proteins identified in the current study upregulated in the SIRD and MDH cluster are generally associated with lower eGFR levels and higher risk for CVD and conversely those downregulated in these two clusters are associated with higher eGFR levels and lower CVD risk. An explanation may be that individuals in the SIRD and MDH cluster are generally older compared to the other three clusters. These results also further confirm the added value of adding HDL to the clustering as the MOD and MD cluster were much more alike than MD and MDH. The proteins upregulated in the MD and MOD cluster were associated with higher eGFR levels and lower CVD risk.

The strengths of the current study include the large number of individuals, the use of multiple cohorts and the use of multiple molecular layers to characterise the clusters. A limitation is that the identified markers are measured in plasma and as such they cannot be directly linked to specific metabolic tissues. Second, whilst we adjusted models for the characteristic of that cluster to identify markers that were not simply proxies of the clinical features that defined the cluster we cannot estimate whether we were able to fully adjust for that characteristic. Third, in the current study we compared the levels of molecular measures between individuals with type 2 diabetes and not relative to healthy controls. We can therefore not infer which cluster would be most close to the general population. Fourth, we use a validated quantitative method to measure metabolites that have previously been linked to diabetes, but the limitation of this targeted method is that other metabolites are not measured. As such, we may have missed metabolites with differential levels across clusters. As such, we may have missed metabolites with differential levels across clusters. Finally, the cohorts used are mainly comprised of people of European descent and these results may not be generalizable to other populations.

CONCLUSION

In the current study, clusters were identified in three cohorts, based on five different clinical characteristics. The underlying molecular signatures of each cluster were markedly different (Fig. 4) suggesting different underlying etiopathological processes. As expected, the identified molecular signatures reflected the underlying phenotype to some extent, but often remained associated after adjustment. Importantly, our study provides important new granularity on the likely molecular processes involved in diabetes pathology in each of the diabetes subgroups.

REFERENCES

1. McCarthy MI: Painting a new picture of personalised medicine for diabetes. Diabetologia 2017;60:793-799
2. Pearson ER: Type 2 diabetes: a multifaceted disease. Diabetologia 2019;62:1107-1112
3. Ahlqvist E, Storm P, Karajamaki A, Martinell M, Dorkhan M, Carlsson A, Vikman P, Prasad RB, Aly DM, Almgren P, Wessman Y, Shaat N, Spegel P, Mulder H, Lindholm E, Melander O, Hansson O, Malmqvist U, Lernmark A, Lahti K, Forsen T, Tuomi T, Rosengren AH, Groop L: Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. The Lancet Diabetes & Endocrinology 2018;6:361–369
4. Slieker RC, Donnelly LA, Fitipaldi H, Bouland GA, Giordano GN, Åkerlund M, Gerl MJ, Ahlqvist E, Ali A, Dragan I, Festa A, Hansen MK, Mansour Aly D, Kim M, Kuznetsov D, Mehl F, Klose C, Simons K, Pavo I, Pullen TJ, Suvitaival T, Wretlind A, Rossing P, Lyssenko V, Legido Quigley C, Groop L, Thorens B, Franks PW, Ibberson M, Rutter GA, Beulens JWJ, 't Hart LM, Pearson ER: Replication and cross-validation of T2D subtypes based on clinical variables: an IMI-RHAPSODY study. medRxiv 2020;
5. Dennis JM, Shields BM, Henley WE, Jones AG, Hattersley ATJTLD, Endocrinology: Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. 2019;7:442-451
6. Zaharia OP, Strassburger K, Strom A, Bönhof GJ, Karusheva Y, Antoniou S, Bódis K, Markgraf DF, Burkart V, Müssig K, Hwang J-H, Asplund O, Groop L, Ahlqvist E, Seissler J, Nawroth P, Kopf S, Schmid SM, Stumvoll M, Pfeiffer AFH, Kabisch S, Tselmin S, Häring HU, Ziegler D, Kuss O, Szendroedi J, Roden M, German Diabetes Study G: Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. The lancet Diabetes & endocrinology 2019;7:684-694
7. Safai N, Ali A, Rossing P, Ridderstråle M: Stratification of type 2 diabetes based on routine clinical markers. Diabetes research and clinical practice 2018;141:275-283
8. Aly DM, Dwivedi OP, Prasad RB, Käräjämäki A, Hjort R, Åkerlund M, Mahajan A, Udler MS, Florez JC, McCarthy MI, Brosnan J, Melander O, Carlsson S, Hansson O, Tuomi T, Groop L, Ahlqvist E: Aetiological differences between novel subtypes of diabetes derived from genetic associations. medRxiv 2020:2020.2009.2029.20203935
9. van der Heijden AA, Rauh SP, Dekker JM, Beulens JW, Elders P, M't Hart L, Rutters F, van Leeuwen N, Nijpels G: The Hoorn Diabetes Care System (DCS) cohort. A prospective cohort of persons with type 2 diabetes treated in primary care in the Netherlands. BMJ open 2017;7:e015599
10. Hebert HL, Shepherd B, Milburn K, Veluchamy A, Meng W, Carr F, Donnelly LA, Tavendale R, Leese G, Colhoun HM, Dow E, Morris AD, Doney AS, Lang CC, Pearson ER,

Smith BH, Palmer CNA: Cohort Profile: Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS). Int J Epidemiol 2018;47:380-381j
11. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N, Cook JP, Schmidt EM, Wuttke M, Sarnowski C, Mägi R, Nano J, Gieger C, Trompet S, Lecoeur C, Preuss MH, Prins BP, Guo X, Bielak LF, Below JE, Bowden DW, Chambers JC, Kim YJ, Ng MCY, Petty LE, Sim X, Zhang W, Bennett AJ, Bork-Jensen J, Brummett CM, Canouil M, Ec Kardt KU, Fischer K, Kardia SLR, Kronenberg F, Läll K, Liu CT, Locke AE, Luan J, Ntalla I, Nylander V, Schönherr S, Schurmann C, Yengo L, Bottinger EP, Brandslund I, Christensen C, Dedoussis G, Florez JC, Ford I, Franco OH, Frayling TM, Giedraitis V, Hackinger S, Hattersley AT, Herder C, Ikram MA, Ingelsson M, Jørgensen ME, Jørgensen T, Kriebel J, Kuusisto J, Ligthart S, Lindgren CM, Linneberg A, Lyssenko V, Mamakou V, Meitinger T, Mohlke KL, Morris AD, Nadkarni G, Pankow JS, Peters A, Sattar N, Stančáková A, Strauch K, Taylor KD, Thorand B, Thorleifsson G, Thorsteinsdottir U, Tuomilehto J, Witte DR, Dupuis J, Peyser PA, Zeggini E, Loos RJF, Froguel P, Ingelsson E, Lind L, Groop L, Laakso M, Collins FS, Jukema JW, Palmer CNA, Grallert H, Metspalu A, Dehghan A, Köttgen A, Abecasis GR, Meigs JB, Rotter JI, Marchini J, Pedersen O, Hansen T, Langenberg C, Wareham NJ, Stefansson K, Gloyn AL, Morris AP, Boehnke M, McCarthy MI: Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 2018;50:1505-1513
12. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C: Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284-1287
13. Ahonen L, Jantti S, Suvitaival T, Theilade S, Risz C, Kostiainen R, Rossing P, Oresic M, Hyotylainen T: Targeted Clinical Metabolite Profiling Platform for the Stratification of Diabetic Patients. Metabolites 2019;9
14. Surma MA, Herzog R, Vasilj A, Klose C, Christinat N, Morin-Rivron D, Simons K, Masoodi M, Sampaio JL: An automated shotgun lipidomics platform for high throughput, comprehensive, and quantitative analysis of blood plasma intact lipids. Eur J Lipid Sci Technol 2015;117:1540-1549
15. Herzog R, Schwudke D, Schuhmann K, Sampaio JL, Bornstein SR, Schroeder M, Shevchenko A: A novel informatics concept for high-throughput shotgun lipidomics based on the molecular fragmentation query language. Genome Biol 2011;12:R8
16. Aimo L, Liechti R, Hyka-Nouspikel N, Niknejad A, Gleizes A, Gotz L, Kuznetsov D, David FP, van der Goot FG, Riezman H, Bougueleret L, Xenarios I, Bridge A: The SwissLipids knowledgebase for lipid biology. Bioinformatics 2015;31:2860-2866
17. Schwarzer G: meta: An R package for meta-analysis. R news 2007;7:40-45
18. Udler MS, Kim J, von Grotthuss M, Bonàs-Guarch S, Mercader JM, Cole JB, Chiou J, Anderson CD, Boehnke M, Laakso M, Atzmon G, Glaser B, Gaulton K, Flannick J, Getz G, Florez JC: Clustering of Type 2 Diabetes Genetic Loci by Multi-Trait Associations Identifies Disease Mechanisms and Subtypes. 2018:319509
19. Yang J, Brody EN, Murthy AC, Mehler RE, Weiss SJ, DeLisle RK, Ostroff R, Williams SA, Ganz P: Impact of Kidney Function on the Blood Proteome and on Protein Cardiovascular Risk Biomarkers in Patients With Stable Coronary Heart Disease. J Am Heart Assoc 2020;9:e016463
20. Hu W, Wei R, Wang L, Lu J, Liu H, Zhang W: Correlations of MMP-1, MMP-3, and MMP-12 with the degree of atherosclerosis, plaque stability and cardiovascular and cerebrovascular events. Exp Ther Med 2018;15:1994-1998

21. McLennan SV, Kelly DJ, Schache M, Waltham M, Dy V, Langham RG, Yue DK, Gilbert RE: Advanced glycation end products decrease mesangial cell MMP-7: a role in matrix accumulation in diabetic nephropathy? Kidney Int 2007;72:481-488
22. Carstensen M, Herder C, Brunner EJ, Strassburger K, Tabak AG, Roden M, Witte DR: Macrophage inhibitory cytokine-1 is increased in individuals before type 2 diabetes diagnosis but is not an independent predictor of type 2 diabetes: the Whitehall II study. Eur J Endocrinol 2010;162:913-917
23. Beg M, Abdullah N, Thowfeik FS, Altorki NK, McGraw TE: Distinct Akt phosphorylation states are required for insulin regulated Glut4 and Glut1-mediated glucose uptake. Elife 2017;6
24. Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, Yang E, Farrell L, Fox CS, O'Donnell CJ, Carr SA, Vasan RS, Florez JC, Clish CB, Wang TJ, Gerszten RE: Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J Clin Invest 2011;121:1402-1411
25. Huang X, Fu C, Liu W, Liang Y, Li P, Liu Z, Sheng Q, Liu P: Chemerin-induced angiogenesis and adipogenesis in 3 T3-L1 preadipocytes is mediated by lncRNA Meg3 through regulating Dickkopf-3 by sponging miR-217. Toxicol Appl Pharmacol 2019;385:114815
26. Belongie KJ, Ferrannini E, Johnson K, Andrade-Gordon P, Hansen MK, Petrie JR: Identification of novel biomarkers to monitor β-cell function and enable early detection of type 2 diabetes risk. PLoS One 2017;12:e0182932
27. Fiodorenko-Dumas Z, Dumas I, Mastej K, Adamiec R: Physical activity - related changes in ADMA and vWF levels in patients with type 2 diabetes: A preliminary study. Adv Clin Exp Med 2017;26:601-608
28. Yu E, Ruiz-Canela M, Razquin C, Guasch-Ferre M, Toledo E, Wang DD, Papandreou C, Dennis C, Clish C, Liang L, Bullo M, Corella D, Fito M, Gutierrez-Bedmar M, Lapetra J, Estruch R, Ros E, Cofan M, Aros F, Romaguera D, Serra-Majem L, Sorli JV, Salas-Salvado J, Hu FB, Martinez-Gonzalez MA: Changes in arginine are inversely associated with type 2 diabetes: A case-cohort study in the PREDIMED trial. Diabetes Obes Metab 2019;21:397-401
29. Ruiz-Canela M, Guasch-Ferré M, Toledo E, Clish CB, Razquin C, Liang L, Wang DD, Corella D, Estruch R, Hernáez Á, Yu E, Gómez-Gracia E, Zheng Y, Arós F, Romaguera D, Dennis C, Ros E, Lapetra J, Serra-Majem L, Papandreou C, Portoles O, Fitó M, Salas-Salvadó J, Hu FB, Martínez-González MAJD: Plasma branched chain/aromatic amino acids, enriched Mediterranean diet and risk of type 2 diabetes: case-cohort study within the PREDIMED Trial. 2018;61:1560-1571
30. Zheng Y, Ceglarek U, Huang T, Li L, Rood J, Ryan DH, Bray GA, Sacks FM, Schwarzfuchs D, Thiery J, Shai I, Qi L: Weight-loss diets and 2-y changes in circulating amino acids in 2 randomized intervention trials. Am J Clin Nutr 2016;103:505-511
31. Meikle PJ, Wong G, Barlow CK, Weir JM, Greeve MA, MacIntosh GL, Almasy L, Comuzzie AG, Mahaney MC, Kowalczyk AJPo: Plasma lipid profiling shows similar associations with prediabetes and type 2 diabetes. 2013;8:e74341
32. Saltevo J, Laakso M, Jokelainen J, Keinänen-Kiukaanniemi S, Kumpusalo E, Vanhala M: Levels of adiponectin, C-reactive protein and interleukin-1 receptor antagonist are associated with insulin sensitivity: a population-based study. Diabetes/metabolism research and reviews 2008;24:378-383
33. Lotta LA, Scott RA, Sharp SJ, Burgess S, Luan J, Tillin T, Schmidt AF, Imamura F, Stewart ID, Perry JR, Marney L, Koulman A, Karoly ED, Forouhi NG, Sjogren RJ, Naslund E, Zierath JR, Krook A, Savage DB, Griffin JL, Chaturvedi N, Hingorani AD, Khaw KT, Barroso I, McCarthy MI, O'Rahilly S, Wareham NJ, Langenberg C: Genetic Predisposition to

an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis. PLoS Med 2016;13:e1002179
34. Gancheva S, Jelenik T, Alvarez-Hernandez E, Roden M: Interorgan Metabolic Crosstalk in Human Insulin Resistance. Physiol Rev 2018;98:1371-1415
35. Li Y, Soos TJ, Li X, Wu J, Degennaro M, Sun X, Littman DR, Birnbaum MJ, Polakiewicz RD: Protein kinase C Theta inhibits insulin signaling by phosphorylating IRS1 at Ser(1101). The Journal of biological chemistry 2004;279:45304-45307
36. Sylow L, Nielsen IL, Kleinert M, Møller LL, Ploug T, Schjerling P, Bilan PJ, Klip A, Jensen TE, Richter EA: Rac1 governs exercise‐stimulated glucose uptake in skeletal muscle through regulation of GLUT4 translocation in mice. The Journal of Physiology 2016;594:4997-5008
37. Ueda S, Kitazawa S, Ishida K, Nishikawa Y, Matsui M, Matsumoto H, Aoki T, Nozaki S, Takeda T, Tamori Y, Aiba A, Kahn CR, Kataoka T, Satoh T: Crucial role of the small GTPase Rac1 in insulin-stimulated translocation of glucose transporter 4 to the mouse skeletal muscle sarcolemma. FASEB journal : official publication of the Federation of American Societies for Experimental Biology 2010;24:2254-2261
38. Jain R, Jain D, Liu Q, Bartosinska B, Wang J, Schumann D, Kauschke S, Eickelmann P, Piemonti L, Gray NJD: Pharmacological inhibition of Eph receptors enhances glucose-stimulated insulin secretion from mouse and human pancreatic islets. 2013;56:1350-1355
39. Koo BK, Joo SK, Kim D, Bae JM, Park JH, Kim JH, Kim W: Additive effects of PNPLA3 and TM6SF2 on the histological severity of non-alcoholic fatty liver disease. J Gastroenterol Hepatol 2018;33:1277-1285
40. Gruzdeva O, Borodkina D, Uchasova E, Dyleva Y, Barbarash O: Leptin resistance: underlying mechanisms and diagnosis. Diabetes, metabolic syndrome and obesity : targets and therapy 2019;12:191-198
41. Lalia AZ, Lanza IR: Insulin-Sensitizing Effects of Omega-3 Fatty Acids: Lost in Translation? Nutrients 2016;8
42. Oh DY, Talukdar S, Bae EJ, Imamura T, Morinaga H, Fan W, Li P, Lu WJ, Watkins SM, Olefsky JM: GPR120 is an omega-3 fatty acid receptor mediating potent anti-inflammatory and insulin-sensitizing effects. Cell 2010;142:687-698
43. Storlien LH, Kraegen EW, Chisholm DJ, Ford GL, Bruce DG, Pascoe WS: Fish oil prevents insulin resistance induced by high-fat feeding in rats. Science 1987;237:885-888

ACKNOWLEDGEMENTS

We acknowledge the support of the Health Informatics Centre, University of Dundee for managing and supplying the anonymised data.

FUNDING SOURCES

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115881 (RHAPSODY). This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. This work is supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 16.0097-2. The opinions expressed and arguments employed herein do not necessarily reflect the official views of these funding

bodies. ERP was supported by a Wellcome Trust investigator award (102820/Z/13/Z). GAR was supported by a Wellcome Trust Senior Investigator (WT098424AIA) and Investigator Award (212625/Z/18/Z), MRC Programme grants (MR/R022259/1, MR/J0003042/1, MR/L020149/1) and by Diabetes UK (BDA/11/0004210, BDA/15/0005275, BDA 16/0005485) project grants.

AUTHOR CONTRIBUTIONS

RCS, LAD, JWJB, LMTH, ERP designed the study and drafted the manuscript. RCS, LAD, HF, GAB, MA performed the analyses. ID, DK, MI set up a federated node system for data-analysis. RCS, DMA, LAD, HF, EA, AA, MJG, MK, FM, TS, AW, CLQ, MI were involved in the data pre-processing and quality control. GNG, AF, MKH, DMA, IP, TJP, BT, VL, LG, PWF, GAR contributed to the data acquisition and project logistics. MJG, CK, KS generated the Lipotype data. CLQ, AA, PR, AW, TS generated the metabolomics data. All authors contributed to the data interpretation. All authors critically revised the manuscript and approved the final version. RCS and LAD are the guarantors of the work.

CONFLICT OF INTEREST

KS is CEO of Lipotype GmbH. KS and CK are shareholders of Lipotype GmbH. MJG is employee of Lipotype GmbH. GAR has received grant funding and consultancy fees from Sun Pharmaceuticals and Les Laboratoires Servier. MKH is an employee of Janssen Research & Development, LLC. AF and IP are employees of Eli Lilly Regional Operations GmbH.

Table 1 Significant SNPs in each of the clusters

Variant	*Cluster*	*Chr*	*Position*	*Gene*	*Risk allele*	*REF*	*ALT*	*Risk AF in cluster*	*Effect*	*Lower*	*Upper*	*P-value*	*I2*	*Heterogeneity*
rs3802177	SIRD	8	118185025	*SLC30A8*	G	G	A	↓	0.07	0.04	0.10	$2.19 \cdot 10^{-5}$	0.09	0.14
rs10811660	SIRD	9	22134068	*CDKN2A/B*	G	G	A	↓	0.05	0.02	0.07	$8.72 \cdot 10^{-5}$	0.00	0.59
rs7903146	SIRD	10	114758349	*TCF7L2*	T	C	T	↓	-0.10	-0.15	-0.05	$1.59 \cdot 10^{-4}$	0.62	0.02
rs11708067	SIRD	3	123065778	*ADCY5*	A	A	G	↓	0.05	0.02	0.08	$3.76 \cdot 10^{-4}$	0.00	0.31
rs243024	SIRD	2	60583665	*BCL11A*	A	G	A	↓	-0.06	-0.10	-0.03	$6.12 \cdot 10^{-4}$	0.10	0.14
rs1421085	MOD	16	53800954	*FTO*	C	T	C	↑	0.06	0.03	0.09	$3.99 \cdot 10^{-5}$	0.00	0.53
rs10893829	MOD	11	128042575	*ETS1*	T	T	C	↓	0.04	0.02	0.06	$6.62 \cdot 10^{-5}$	0.00	0.35
rs523288	MOD	18	57848369	*MC4R*	T	A	T	↑	0.05	0.02	0.08	$1.54 \cdot 10^{-4}$	0.00	0.23
rs8107974	MOD	19	19388500	*TM6SF2*	T	A	T	↓	-0.04	-0.06	-0.02	$2.60 \cdot 10^{-4}$	0.32	0.09
rs10096633	MD	8	19830921	*LPL*	C	C	T	↑	-0.04	-0.05	-0.02	$7.60 \cdot 10^{-5}$	0.00	0.57
rs10096633	MDH	8	19830921	*LPL*	C	C	T	↓	0.07	0.05	0.09	$1.04 \cdot 10^{-11}$	0.00	0.25

FIGURE LEGENDS

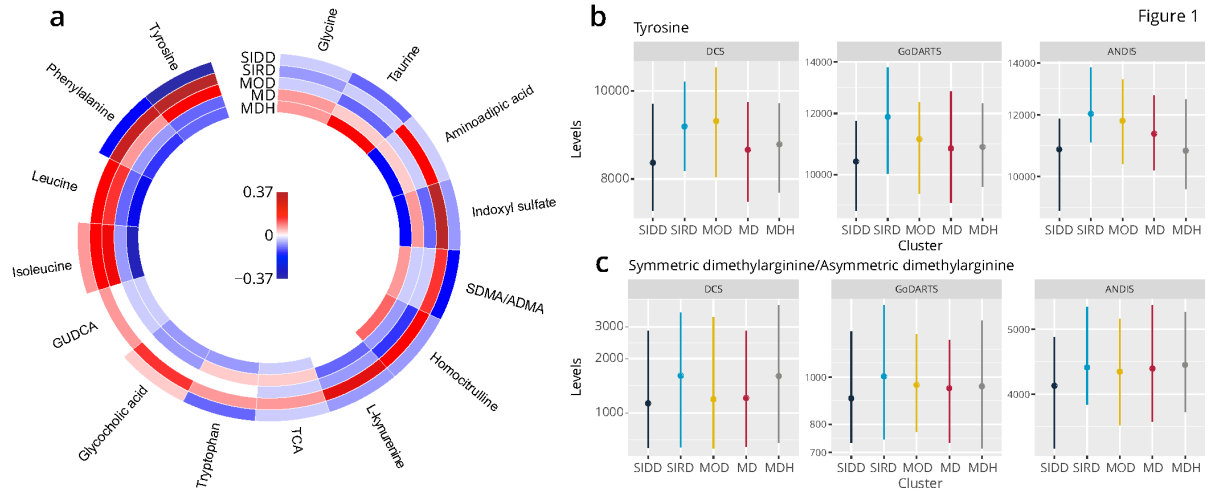

Figure 1 Metabolite levels in the five clusters. a. Change in metabolites levels in each of the clusters versus all others. Colours represent effect size in log SD, with red upregulation and blue downregulation. SDMA/ADMA, Symmetric dimethylarginine/Asymmetric dimethylarginine; TCA, Taurocholic acid; GUDCA, Glycoursodeoxycholic acid. **b.** Levels of tyrosine in DCS, GoDARTS and ANDIS. SIDD and SIRD $P_{FDR} < 0.05$. **c.** Levels of (a) symmetric dimethylarginine. SIDD and SIRD $P_{FDR} \leq 0.05$. Dots represent the median, the vertical line the interquartile range. SIDD, Severe Insulin-Deficit Diabetes; SIRD, Severe Insulin-Resistant Diabetes cluster; MOD, Mild Obesity-related Diabetes; MD, Mild diabetes; MDH, Mild diabetes with high HDL.

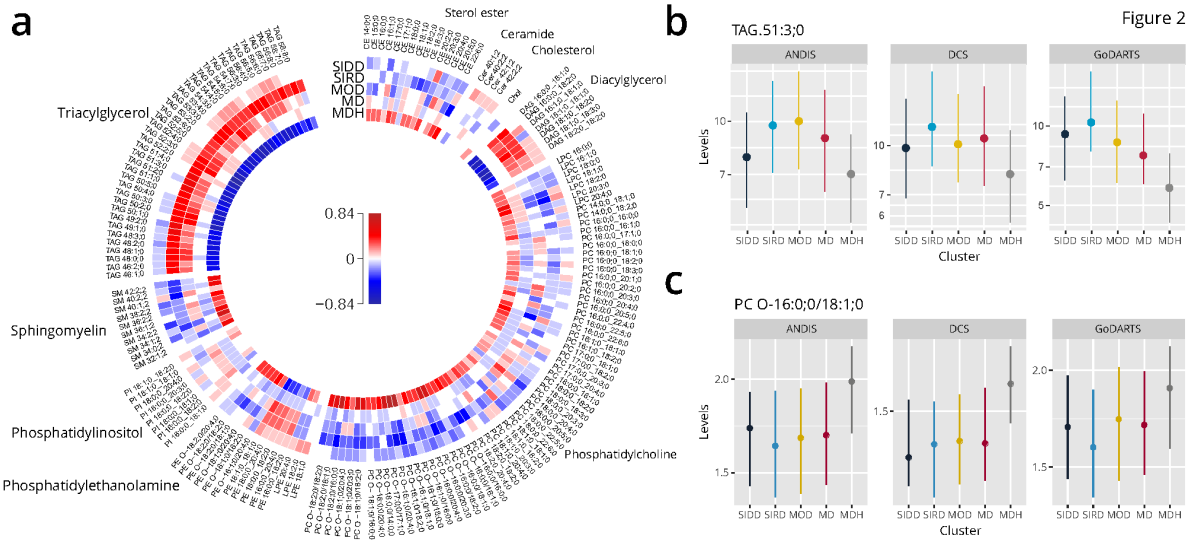

Figure 2. Lipid levels in the five clusters. a. Change in lipid levels in each of the clusters versus all others. Colours represent effect size in log SD, with red upregulation and blue downregulation b**.** Levels of TAG 51:3;0 in DCS, GoDARTS and ANDIS. SIRD, MOD and MDH $P_{FDR} \leq 0.05$. **c.** Levels of PC O-16:0;0/18:1;0. SIRD, MOD and MDH $P_{FDR} \leq 0.05$. Dots

represent the median, the vertical line the interquartile range. SIDD, Severe Insulin-Deficit Diabetes; SIRD, Severe Insulin-Resistant Diabetes cluster; MOD, Mild Obesity-related Diabetes; MD, Mild diabetes; MDH, Mild diabetes with high HDL.

Figure 3

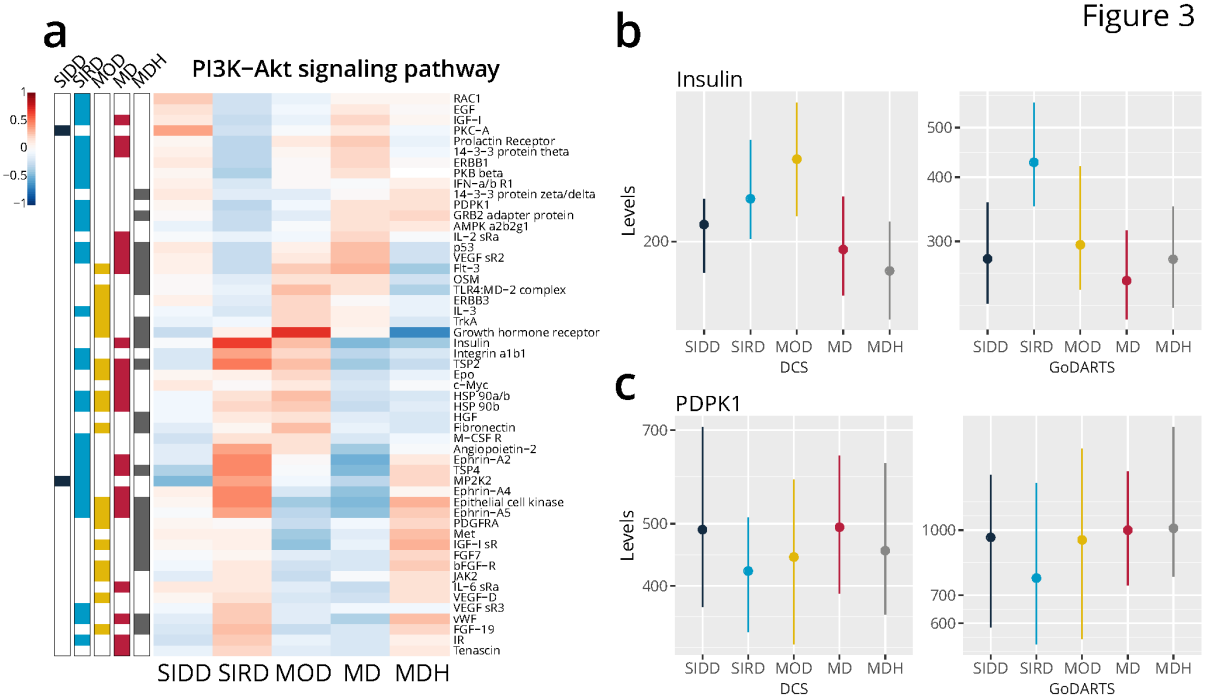

Figure 3 Proteins in the PI3K/Akt pathway in the five clusters. a. Effect sizes of proteins in the PI3K/Akt pathway ($P_{FDR}=1.05 \cdot 10^{-29}$) with red upregulation in the cluster versus all others and blue downregulation. Bars on the left indicate whether proteins are statistically significant in a specific cluster. Dots represent the median, the vertical line the interquartile range. **b.** Levels of insulin in DCS, GoDARTS and ANDIS. MDH $P_{FDR} \leq 0.05$. **c.** Levels of PDPK1. Dots represent the median, the vertical line the interquartile range. SIRD $P_{FDR} \leq 0.05$.

Figure 4

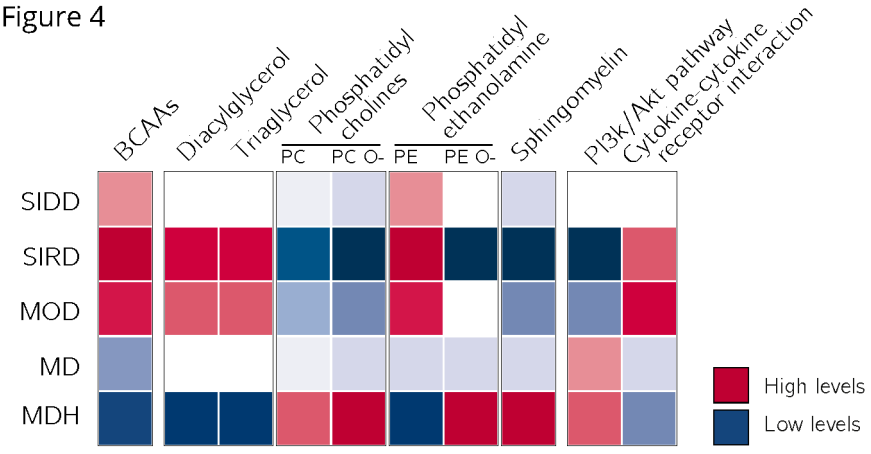

Figure 4 Schematic overview of the results in the current study. BCAAs, DAGs, TAGs, PE were upregulated in SIRD and to a lesser extend MOD, while being downregulated in MDH. PE O-, sphingomyelins and proteins associated with the PI3K/Akt pathway were

downregulated in SIRD. In MOD proteins were found upregulated that have been associated with cytokine-cytokine interaction. SIDD, Severe Insulin-Deficit Diabetes; SIRD, Severe Insulin-Resistant Diabetes cluster; MOD, Mild Obesity-related Diabetes; MD, Mild diabetes; MDH, Mild diabetes with high HDL.